Ridge Regression

Nipun Batra

IIT Gandhinagar

July 29, 2025

In $f(x) = c_0 + c_1 x + c_2 x^2 + \ldots$ it is $\max |c_i|$

Objective:

$$\begin{aligned} \text{Minimise } & (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \\ \text{s.t. } & \boldsymbol{\theta}^T \boldsymbol{\theta} \leq S \end{aligned}$$

Objective:

$$\begin{aligned} \text{Minimise } & (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \\ \text{s.t. } & \boldsymbol{\theta}^T \boldsymbol{\theta} \leq S \end{aligned}$$

This is equivalent to

$$\text{Minimise } (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) + \delta^2 \boldsymbol{\theta}^T \boldsymbol{\theta}$$

$$\text{Minimise } (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})$$
$$\text{s.t. } \boldsymbol{\theta}^T \boldsymbol{\theta} \leq S$$
$$L(\boldsymbol{\theta}, \mu) = (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) + \mu \left(\boldsymbol{\theta}^T \boldsymbol{\theta} - S \right)$$

where, $\mu \geq 0$ (and $\mu = \delta^2$)

$$
\begin{aligned}
\text{Minimise } & (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\theta}) \\
\text{s.t. } & \boldsymbol{\theta}^T\boldsymbol{\theta} \leq S \\
L(\boldsymbol{\theta},\mu) = & (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\theta}) + \mu\left(\boldsymbol{\theta}^T\boldsymbol{\theta} - S\right)
\end{aligned}
$$

where, $\mu \geq 0$ (and $\mu = \delta^2$)

$$\text{Minimise } (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})$$

$$\text{s.t. } \boldsymbol{\theta}^T\boldsymbol{\theta} \leq S$$

$$L(\boldsymbol{\theta},\mu) = (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\theta}) + \mu\left(\boldsymbol{\theta}^T\boldsymbol{\theta} - S\right)$$

where, $\mu \geq 0$ (and $\mu = \delta^2$)

If $\mu = 0$
There is no regularization
No effect on constraint

$$\text{Minimise } \left(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\right)^T \left(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\right)$$
$$\text{s.t. } \boldsymbol{\theta}^T\boldsymbol{\theta} \leq S$$
$$L\left(\boldsymbol{\theta}, \mu\right) = \left(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\right)^T \left(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\right) + \mu\left(\boldsymbol{\theta}^T\boldsymbol{\theta} - S\right)$$

where, $\mu \geq 0$ (and $\mu = \delta^2$)

If $\mu = 0$
There is no regularization
No effect on constraint

$$\text{Minimise } (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})$$
$$\text{s.t. } \boldsymbol{\theta}^T\boldsymbol{\theta} \leq S$$
$$L(\boldsymbol{\theta},\mu) = (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\theta}) + \mu\left(\boldsymbol{\theta}^T\boldsymbol{\theta} - S\right)$$

where, $\mu \geq 0$ (and $\mu = \delta^2$)

If $\mu = 0$
There is no regularization
No effect on constraint

If $\mu \neq 0$
$\implies \boldsymbol{\theta}^T\boldsymbol{\theta} - S = 0$

Are θ_i all zero for high μ?

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{20}\begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{X} + \mu\mathbf{I} = \begin{bmatrix} 6 & 10 \\ 10 & 32 \end{bmatrix}$$

$$(\mathbf{X}^T\mathbf{X} + \mu\mathbf{I})^{-1} = \frac{1}{92} \begin{bmatrix} 32 & -10 \\ -10 & 6 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

Another interpretation of “regularisation”

- Contrast with update equation for unregularised regression:

- Contrast with update equation for unregularised regression:
- $\boldsymbol{\theta} = \underbrace{\boldsymbol{\theta}}_{\text{No Shrinking } \boldsymbol{\theta}} - \alpha(-2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\theta})$